

Unsound Inferences Make Proofs Shorter*

Juan P. Aguilera¹ and Matthias Baaz¹

¹*Vienna University of Technology, Austria.*

Abstract

We give examples of calculi that extend Gentzen's sequent calculus LK by unsound quantifier inferences in such a way that (i) derivations lead only to true sequents, and (ii) proofs therein are non-elementarily shorter than LK-proofs.

1 Introduction

Consider the following argument:

1. That Kurt Gödel is Austrian entails that Kurt Gödel is Austrian.
2. Hence, that Kurt Gödel is Austrian entails that everyone is Austrian.
3. That is, it is a fact that if Kurt Gödel is Austrian, then all people are Austrian.
4. Therefore, there exists a person such that, if that person is Austrian, then all people are Austrian.

The argument is an instance of the following proof schema:

$$\frac{\frac{\frac{A(a) \vdash A(a)}{A(a) \vdash \forall y A(y)}}{\vdash A(a) \rightarrow \forall y A(y)}}{\vdash \exists x (A(x) \rightarrow \forall y A(y))} \quad (1)$$

The study of proofs is motivated by our desire for deductive reasoning to be *correct*, i.e., we wish that it be such that the procedures involved derive only true conclusions. The traditional way of ensuring this involves restricting proofs to those satisfying the two following properties:

- Inferences are *sound*¹, i.e., only true conclusions result from true premises.

⁰Partially supported by FWF grants P-26976-N25, I-1897-N25, I-2671-N35, and W1255-N23.

¹Soundness is traditionally applied to derivations or logical systems. In this paper, we distinguish 'soundness' from 'correctness.' See below.

- Derivations are *hereditary*, i.e., initial segments of proofs are proofs themselves.

We challenge this practice. We give examples of correct logical calculi augmented with unsound rules. In our calculi, inferences such as (1) will be allowed.

In our current treatment, the correctness of proofs can only be gauged by considering them in their entirety, as forfeiting soundness while maintaining correctness necessarily violates hereditariness. This is not unlike other aspects of reasoning that are traditionally not reflected in formal derivations. For example, hereditariness is not compatible with proofs by contradiction—one would not take as correct an initial segment of a proof obtained by interrupting it after an assumption directed towards a contradiction had been made, but before the contradiction is reached.

To illustrate this, we note that the following subproof of the opening example should not be allowed as a proof in the system:

1. That Kurt Gödel is Austrian entails that Kurt Gödel is Austrian.
2. Hence, that Kurt Gödel is Austrian entails that everyone is Austrian.

A key factor in ensuring that our proof of 4. above is correct is that in reality it says nothing about Kurt Gödel. His name might well have been replaced with any other with no effect on the derivation; we call this condition *substitutability*.

Traditional proofs can be taken without loss of generality to be *regular*, i.e., any variable which is the *characteristic variable* of a quantifier inference can be assumed to appear only before then. Forfeiting this condition makes it possible for us to accept the following proof:

$$\begin{array}{c}
 \frac{A(a) \vdash A(a)}{A(a) \vdash \forall x A(x)} \quad \frac{A(f(a)) \vdash A(f(a))}{\forall x A(x) \vdash A(f(a))} \\
 \hline
 \frac{A(a) \vdash A(f(a))}{\vdash A(a) \rightarrow A(f(a))} \\
 \hline
 \vdash \exists x (A(x) \rightarrow A(f(x))) \tag{2}
 \end{array}$$

Note that in this example, however, a is only the characteristic variable of one quantifier inference.² We call this *weak regularity*. Our main result is that a sequent calculus augmented with unsound quantifier inferences satisfying substitutability, weak regularity (or even a further weakening thereof), and a technical *side-variable condition* is correct and yields non-elementarily-shorter cut-free proofs than the standard proof systems.

2 Quantifier Inferences

We consider derivations in sequent calculi. Sequents are expressions of the form $\Gamma \vdash \Delta$, where Γ and Δ represent collections of formulae. The sequent $\Gamma \vdash \Delta$ is

²See Section 2.

usually interpreted as 'if all of Γ hold, then at least one of the formulae in Δ holds.' Inferences are expressions of the form

$$\frac{\Gamma \vdash \Delta}{\Gamma' \vdash \Delta'}$$

which are to be interpreted as 'the sequent $\Gamma' \vdash \Delta'$ follows from the sequent $\Gamma \vdash \Delta$.' Derivations are trees whose leaves are axioms and whose non-leaf nodes are inferences; they are to be interpreted as *proofs* of the root sequent.

A formula is *critical* in an inference if it appears in the conclusion but not in the premise. An inference is a *quantifier inference* if the (only) critical formula has a quantifier as its outermost logical symbol. Normally, quantifier inferences consist of substituting a formula³ $Qx A(x)$ for an instance thereof in a sequent, e.g., as in

$$\frac{\Gamma \vdash \Delta, A(a)}{\Gamma \vdash \Delta, \forall x A(x)}$$

The *polarity* of a quantifier is defined as follows: a quantifier is *positive* in a formula rewritten without implication symbols if it is on the right-hand side (resp. left-hand side) of a sequent and under the scope of an even (resp. odd) number of negation symbols, and *negative* otherwise. A quantifier is *strong* if it is positive and on the right-hand side of a sequent or weak and on the left-hand side of a sequent; it is *weak* otherwise.

If an inference yields a strongly-quantified formula $Qx A(x)$ from $A(a)$, where a is a free variable, we say that a is the *characteristic variable* of the inference. We will denote free variables by letters a, b, c, \dots (in contrast, we denote bound variables by x, y, z, \dots). We denote closed terms by t, s, r, \dots and variants thereof.

Let π be any derivation. We say b is a *side variable* of a in π (written $a <_{\pi} b$, or simply $a < b$) if π contains a strong-quantifier inference of the form:

$$\frac{\Gamma \vdash \Delta, A(a, b, \vec{c})}{\Gamma \vdash \Delta, \forall x A(x, b, \vec{c})}$$

or of the form:

$$\frac{A(a, b, \vec{c}), \Gamma \vdash \Delta}{\exists x A(x, b, \vec{c}), \Gamma \vdash \Delta}$$

Definition 2.1 (Suitable quantifier inference). We say a quantifier inference is *suitable for a proof* π if either it is a weak-quantifier inference, or it satisfies the following three conditions:

- (substitutability) the characteristic variable does not appear in the root of π .
- (side-variable condition) the relation $<_{\pi}$ is acyclic.
- (weak regularity) the characteristic variable is not the characteristic variable of another strong-quantifier inference in π .

³We will frequently use Q to denote any unspecified quantifier.

We will work with various sequent calculi extending *Propositional* LK, the sequent calculus whose only axiom is $A \vdash A$ —with A atomic—and whose rules are:

Structural rules:

$$\begin{array}{c}
\frac{\Gamma \vdash \Delta}{\Gamma, A \vdash \Delta} \text{ } WL \\
\frac{\Gamma, A, A \vdash \Delta}{\Gamma, A \vdash \Delta} \text{ } CL \\
\frac{\Gamma_1, A, B, \Gamma_2 \vdash \Delta}{\Gamma_1, B, A, \Gamma_2 \vdash \Delta} \text{ } EL \\
\frac{\Gamma \vdash \Delta, A \quad A, \Gamma \vdash \Delta}{\Gamma \vdash \Delta} \text{ } Cut
\end{array}
\qquad
\begin{array}{c}
\frac{\Gamma \vdash \Delta}{\Gamma \vdash \Delta, A} \text{ } WR \\
\frac{\Gamma \vdash \Delta, A, A}{\Gamma \vdash \Delta, A} \text{ } CR \\
\frac{\Gamma \vdash \Delta_1, A, B, \Delta_2}{\Gamma \vdash \Delta_1, B, A, \Delta_2} \text{ } ER
\end{array}$$

Logical rules:

$$\begin{array}{c}
\frac{\Gamma, A \vdash \Delta}{\Gamma, A \wedge B \vdash \Delta} \wedge_1 L \\
\frac{\Gamma, B \vdash \Delta}{\Gamma, A \wedge B \vdash \Delta} \wedge_2 L \\
\frac{\Gamma, A \vdash \Delta \quad \Gamma, B \vdash \Delta}{\Gamma, A \vee B \vdash \Delta} \vee L \\
\frac{\Gamma \vdash \Delta, A \quad \Gamma, B \vdash \Delta}{\Gamma, A \rightarrow B \vdash \Delta} \rightarrow L \\
\frac{\Gamma \vdash A, \Delta}{\Gamma, \neg A \vdash \Delta} \neg L
\end{array}
\qquad
\begin{array}{c}
\frac{\Gamma \vdash \Delta, A}{\Gamma \vdash \Delta, A \vee B} \vee_1 R \\
\frac{\Gamma \vdash \Delta, B}{\Gamma \vdash \Delta, A \vee B} \vee_2 R \\
\frac{\Gamma \vdash \Delta, A \quad \Gamma \vdash \Delta, B}{\Gamma \vdash \Delta, A \wedge B} \wedge R \\
\frac{\Gamma, A \vdash \Delta, B}{\Gamma \vdash \Delta, A \rightarrow B} \rightarrow R \\
\frac{\Gamma, A \vdash \Delta}{\Gamma \vdash \Delta, \neg A} \neg R
\end{array}$$

The names of the structural rules stand, respectively, for ‘weakening,’ ‘contraction,’ and ‘exchange.’ First-order LK is the extension of Propositional LK obtained by adding quantifier inferences:

$$\begin{array}{c}
\frac{\Gamma, A(t) \vdash \Delta}{\Gamma, \forall x A(x) \vdash \Delta} \forall L \\
\frac{\Gamma, A(a) \vdash \Delta}{\Gamma, \exists x A(x) \vdash \Delta} \exists L
\end{array}
\qquad
\begin{array}{c}
\frac{\Gamma \vdash \Delta, A(t)}{\Gamma \vdash \Delta, \exists x A(x)} \exists R \\
\frac{\Gamma \vdash \Delta, A(a)}{\Gamma \vdash \Delta, \forall x A(x)} \forall R
\end{array}$$

with the only restriction that in the inferences $\forall R$ and $\exists L$, the variable a is not allowed to appear in the conclusion. In particular, note that every quantifier inference is suitable for every LK-proof. Gentzen’s famous *Cut-Elimination Theorem* states that the cut-rule is redundant in both propositional and first-order LK.

Definition 2.2 (LK⁺). The calculus LK⁺ is defined as LK, except that we instead allow all weak and strong quantifier inferences with the proviso that they be suitable for the proof.

A further weakening of the characteristic variable condition gives rise to the notion of weak suitability:

Definition 2.3 (Weakly suitable quantifier inference). A quantifier inference is *weakly suitable for a proof* π if either it is a weak-quantifier inference or it satisfies substitutability, the side-variable condition, and:

- (very weak regularity) whenever the characteristic variable is also the characteristic variable of another strong-quantifier inference in π , then it has the same critical formula.

Definition 2.4 (LK^{++}). The calculus LK^{++} is the extension of LK that results from allowing all weakly suitable quantifier inferences.

It is easy to find examples of sequents that are more easily provable in LK^{++} than in LK :

Example 2.5. *The sequent*

$$\forall x A(x) \rightarrow B \vdash \exists x (A(x) \rightarrow B)$$

is provable in LK :

$$\frac{\frac{\frac{\frac{\frac{\frac{A(a) \vdash A(a)}{A(a) \vdash A(a), B}}{\vdash A(a), A(a) \rightarrow B}}{\vdash A(a), \exists x (A(x) \rightarrow B)}}{\vdash \exists x (A(x) \rightarrow B), A(a)}}{\vdash \exists x (A(x) \rightarrow B), \forall x A(x)} \quad B \vdash B}{\forall x A(x) \rightarrow B \vdash \exists x (A(x) \rightarrow B), B}}{\forall x A(x) \rightarrow B, A(b) \vdash \exists x (A(x) \rightarrow B), B}}{\frac{\forall x A(x) \rightarrow B \vdash \exists x (A(x) \rightarrow B), A(b) \rightarrow B}{\forall x A(x) \rightarrow B \vdash \exists x (A(x) \rightarrow B), \exists x (A(x) \rightarrow B)}}{\forall x A(x) \rightarrow B \vdash \exists x (A(x) \rightarrow B)}}$$

However, one can find a shorter LK^{++} -proof:

$$\frac{\frac{\frac{A(a) \vdash A(a)}{A(a) \vdash \forall x A(x)} \quad B \vdash B}{A(a), \forall x A(x) \rightarrow B \vdash B}}{\forall x A(x) \rightarrow B, A(a) \vdash B}}{\forall x A(x) \rightarrow B \vdash A(a) \rightarrow B}}{\forall x A(x) \rightarrow B \vdash \exists x (A(x) \rightarrow B)}}$$

□

Recall that a function on the natural numbers is *exponential* if it can be defined by a quantifier-free formula from $+$, \cdot , and the function $x \mapsto 2^x$. A function on the natural numbers is *elementary* if it can be defined by a quantifier-free formula from $+$, \cdot , $x \mapsto 2^x$, and the function

$$n \mapsto \underbrace{2^{2^{\cdot^{\cdot^2}}}}_{n \text{ times}}.$$

By independent results of R. Statman [7] (which were formalized in [2]) and of V. P. Orevkov [6], the sizes of the smallest cut-free LK-proofs of sequents of length n are not bounded by any elementary function on n . Our main theorem is that cut-free LK⁺-proofs are non-elementarily shorter than cut-free LK-proofs:

Theorem 2.6. *There is no elementary function bounding the length of the shortest cut-free LK-proof of a formula in terms of its shortest cut-free LK⁺-proof.*

An immediate consequence is the following:

Corollary 2.7. *There is no elementary function bounding the length of the shortest cut-free LK-proof of a formula in terms of its shortest cut-free LK⁺⁺-proof.*

We postpone the proof of Theorem 2.6 until Section 4. First, we consider the question of the correctness of LK⁺⁺.

3 Correctness

In this section, we prove the following:

Theorem 3.1. *If a sequent is derivable in LK⁺⁺, then it is already derivable in LK.*

As before, the following is an immediate consequence:

Corollary 3.2. *If a sequent is derivable in LK⁺⁺, then it is already derivable in LK.*

A few remarks on terminology are in order. Below, if A is a subformula of B , we say that B is a *context* of A . We generally write $\kappa[A]$ to emphasize that A occurs in a context $\kappa[\cdot]$; they should be thought as syntactical operators $A \mapsto \kappa[A]$.

The Skolemization of a first-order formula is defined by replacing every strongly quantified variable y with a new function symbol $f_y(x_1, \dots, x_n)$, where x_1, \dots, x_n are the weakly quantified variables such that Qy appears in the scope of their quantifiers. We write $sk(A)$ for the Skolemization of A . The Skolemization of a sequent

$$A_0, \dots, A_m \vdash A_{m+1}, \dots, A_{m+n}$$

is defined as:

$$sk(A_0), \dots, sk(A_m) \vdash sk(A_{m+1}), \dots, sk(A_{m+n}).$$

Skolemization preserves validity, whence by completeness, a sequent is derivable if and only if its Skolemization is. Herbrand's theorem states that a sequent containing only weak quantifiers:

$$A_0, \dots, A_m \vdash A_{m+1}, \dots, A_{m+n} \quad (3)$$

is LK-provable if and only if for each subformula $Qx B(x)$ of A_j such that A_j is of the form:

$$\kappa[Qx B(x)] \quad (4)$$

there are finitely many instances $B_j^i = B(t_j^i)$ of $B(x)$ such that

$$A_0^*, \dots, A_m^* \vdash A_{m+1}^*, \dots, A_{m+n}^* \quad (5)$$

is a propositional tautology, where A_j^* is the result of iteratively replacing each formula (4) with its Herbrand disjunction:

$$\kappa[\bigwedge_i B_j^i],$$

if $j \leq m$, or

$$\kappa[\bigvee_i B_j^i],$$

if $m < j$. If so, we refer to (5) as the *Herbrand sequent* of (3). In particular, if each A_j is prenex of the form:

$$Qx_0 \dots Qx_k A,$$

then there are finitely many instances A_j^i of A such that:

$$\bigwedge_i A_0^i, \dots, \bigwedge_i A_m^i \vdash \bigvee_i A_{m+1}^i, \dots, \bigvee_i A_{m+n}^i$$

is a propositional tautology.

Definition 3.3. The quantifier-free language \mathcal{L}_ε is obtained by removing symbols \exists and \forall from the language of first-order logic and adding symbols ε and τ . Formulae and terms are simultaneously defined by induction in a way that the following clauses are satisfied. We leave the precise recursion to the reader.

1. constants and free variables are terms;
2. if t_1, \dots, t_n are terms and f is an n -ary function symbol, then $f(t_1, \dots, t_n)$ is a term;
3. if $A(t)$ is a formula, t is a term, and x is a bound variable, then $\varepsilon_x A(x)$ and $\tau_x A(x)$ are terms.

4. if t_0, \dots, t_n are terms, then $A(t_0, \dots, t_n)$ is a formula, n -ary predicate symbol A ;
5. if $A(a)$ is a formula and x is a bound variable, then $\forall x A(x)$ and $\exists x A(x)$ are formulae;
6. if A and B are formulae, then $A \wedge B$, $A \vee B$, $\neg A$, and $A \rightarrow B$ are formulae.

(A sequent-calculus reformulation of) Hilbert's ε -calculus (see [4]) LK^ε is obtained by adding to Propositional LK the following rules:

$$\frac{\Gamma, A(t) \vdash \Delta}{\Gamma, A(\tau_x A(x)) \vdash \Delta} \tau \qquad \frac{\Gamma \vdash \Delta, A(t)}{\Gamma \vdash \Delta, A(\varepsilon_x A(x))} \varepsilon$$

The term $\varepsilon_x A$ is to be understood as a 'generic' object satisfying property A . The dual term $\tau_x A$ is added for symmetry; it is in fact redundant—one can define $A(\tau_x A) = A(\varepsilon_x (\neg A))$. The *standard translation* of first-order logic into epsilon calculus is defined by mapping everything to itself, except for quantified formulae: suppose A' is the standard translation of A , then:

$$\exists x A(x) \mapsto A'(\varepsilon_x A'(x)), \qquad \forall x A(x) \mapsto A'(\tau_x A'(x))$$

we write $tr(A) = A'$ to mean that A' is the standard translation of a first-order formula A .

In analogy with the first-order terminology, a term $\varepsilon_x A(x)$ is *strong* in a formula rewritten without implication symbols if it is on the left-hand side and under the scope of an even number of negation symbols, or on the right-hand side and under the scope of an odd number of negation symbols, and *weak* otherwise. Similarly, a term $\tau_x A(x)$ is strong in a formula rewritten without implication symbols if it is on the right-hand side and under the scope of an even number of negation symbols, or on the left-hand side and under the scope of an odd number of negation symbols, and weak otherwise.

We say an \mathcal{L}_ε -sequent is in *weak form* if all occurrences of ε and τ are weak. The analog of Herbrand's theorem for epsilon calculus is called the *Extended First Epsilon Theorem* (see [5, Theorem 16]). It states the following: let

$$A_0, \dots, A_m \vdash A_{m+1}, \dots, A_{m+n}$$

be an LK^ε -provable sequent in weak form and let \vec{t}^i be all the terms of the form $\tau_x A(x)$ or $\varepsilon_x A(x)$ in A_i . Then there exist finitely-many tuples of terms \vec{t}^i_j without any occurrence of the symbols τ or ε such that if A_j^i is the result of substituting the terms \vec{t}^i_j for \vec{t}^i in A_i , then the Herbrand sequent

$$\bigwedge_i A_0^i, \dots, \bigwedge_i A_m^i \vdash \bigvee_i A_{m+1}^i, \dots, \bigvee_i A_{m+n}^i$$

is a propositional tautology. If a first-order sequent is prenex and contains only weak quantifiers, then the Herbrand sequent obtained via Herbrand's theorem and the one obtained by applying the Extended First Epsilon Theorem to its standard translation are the same. ε -calculus is correct in the following sense:

Proposition 3.4. *A Skolemized sequent is LK^ε -derivable whenever its standard translation is LK^ε -derivable.*

Proof. Suppose π is an LK^ε -proof of a sequent $\Gamma_0 \vdash \Delta_0$ that is the standard translation of a Skolemized sequent in first-order logic, say $\Sigma_0 \vdash \Xi_0$. This sequent need not be prenex. If not, we extend π to obtain an LK^ε -proof of a sequent $\Gamma \vdash \Delta$ that is the standard translation of a Skolemized and prenex sequent in first-order logic: let $\kappa[Qx A(x)]$ be the first (i.e., leftmost) formula in the sequent that is not the translation of a prenex formula and Q be the first occurrence of a quantifier witnessing this. Assume, without loss of generality that Q is \exists .

Hence, $\text{tr}(\kappa[\exists x A(x)])$ is of the form $\kappa'[A'(\varepsilon_x A'(x))]$ for some context κ' and some formula A' (which might be different from κ and A if these contain further weak quantifiers), so that the root of π has the form:

$$\Gamma \vdash \Delta_0^0, \kappa'[A'(\varepsilon_x A'(x))], \Delta_0^1.$$

Letting

$$B(\cdot) = \kappa'[A'(\cdot)],$$

and

$$t = \varepsilon_x A'(x),$$

the root of π has the form:

$$\Gamma \vdash \Delta_0^0, B(t), \Delta_0^1.$$

Extend π with an inference ε applied to $B(t)$:

$$\frac{\Gamma \vdash \Delta_0^0, B(t), \Delta_0^1}{\Gamma \vdash \Delta_0^0, B(\varepsilon_y B(y)), \Delta_0^1} \varepsilon$$

Notice that:

$$\begin{aligned} B(\varepsilon_y B(y)) &= B(\varepsilon_y \kappa'[A'(y)]) \\ &= \kappa'[A'(\varepsilon_y \kappa'[A'(y)])] \\ &= \text{tr}(\exists y \kappa[A(y)]). \end{aligned}$$

Therefore, after one inference, this quantifier is no longer a witness to the fact that the final sequent in the LK^ε -proof is not the standard translation of a non-prenex sequent. By repeating this procedure for each witness, we arrive at an extended proof of the desired sequent $\Gamma \vdash \Delta$. Note that throughout the transformation we do not alter the structure of the first-order sequent except for the position of weak quantifiers.

We apply the Extended First Epsilon Theorem to see that the Herbrand sequent S of $\Gamma \vdash \Delta$ is a propositional tautology. But this is also the sequent

obtained by applying Herbrand's theorem to the sequent whose translation is $\Gamma \vdash \Delta$, say $\Sigma \vdash \Xi$.

Recall that the Herbrand sequent results by replacing weak quantifiers with finite disjunctions (on the right-hand side) or conjunctions (on the left-hand side) of instances of the replaced sequent. The key point is that in passing from $\Sigma_0 \vdash \Xi_0$ to $\Sigma \vdash \Xi$, we do not change the structure of any formula, except for the position of quantifiers (without changing their polarity, as all quantifiers are weak). Hence, all instances in S of each formula in Σ or Ξ are also instances of the corresponding formula in Σ_0 or Ξ_0 , so that S is the Herbrand sequent of $\Sigma_0 \vdash \Xi_0$. Therefore, by Herbrand's theorem, $\Sigma_0 \vdash \Xi_0$ is LK-derivable. \square

Proposition 3.5. *If a Skolemized sequent is LK^{++} -derivable, then its standard translation is LK^ε -derivable.*

Proof. Let π be a LK^{++} -proof of a sequent. We transform it into an LK^ε -proof as follows:

Go downwards through each branch of the proof from left to right. Every time we reach a strong quantifier inference, say,

$$\frac{\Gamma \vdash \Delta, A(a, \vec{b})}{\Gamma \vdash \Delta, \forall x A(x, \vec{b})}$$

omit the inference and all other inferences

$$\frac{\Sigma \vdash \Xi, A(a, \vec{b})}{\Sigma \vdash \Xi, \forall x A(x, \vec{b})}$$

and modify the proof by

$$\text{substituting } \tau_x A(x, \vec{b}) \text{ for } a \tag{6}$$

throughout the proof. This is always possible: the side-variable condition ensures that the term $\tau_x A(x, \vec{b})$ makes sense.⁴ Moreover, we can freely replace a since, by very weak regularity, it is the characteristic variable only of inferences with critical formula $\forall x A(x, \vec{b})$. Every time we reach a weak quantifier inference, say,

$$\frac{\Gamma \vdash \Delta, A(t, \vec{b})}{\Gamma \vdash \Delta, \exists x A(x, \vec{b})} \tag{7}$$

substitute it with an inference

$$\frac{\Gamma \vdash \Delta, A(t, \vec{b})}{\Gamma \vdash \Delta, A(\varepsilon_x A(x, \vec{b}), \vec{b})} \tag{8}$$

⁴Suppose we had not made this assumption and a were substituted by $\tau_x A(x, b)$. This leaves the possibility that later on in the derivation there is a strong-quantifier inference of $\forall x B(x, a)$ from $B(b, a)$, in which case we would not be able to substitute b for a . This is the only part of the proof where we make use of the side-variable condition.

Note that this is still possible after any substitutions of the form (6).

This transformation results in a correct proof: any inference not considered thus far is either a structural rule or a propositional logical rule and remains valid. Finally, note that the resulting formula is in fact the standard translation of the sequent whose proof is π . This follows from the fact that we substitute all inferences (7) with inferences (8) and the fact that all occurrences of any free variable that was substituted by (6) do not figure in the end sequent, by substitutability. \square

Proposition 3.6. *If $\Gamma \vdash \Delta$ is LK^{++} -derivable, then its Skolemization is derivable in quadratically many steps using additional cuts.*

Proof. Let $\Gamma \vdash \Delta$ be any LK^{++} -derivable sequent. For each formula A in Δ , $A \vdash sk(A)$ is already LK-derivable, it follows that, by using one cut per formula, we can derive $\Gamma \vdash sk(\Delta)$ in LK^{++} . Similarly, $sk(A) \vdash A$ is already LK-derivable. Hence, by also using one cut for each formula in Γ , we can derive $sk(\Gamma) \vdash sk(\Delta)$. \square

Correctness readily follows:

Proof of Theorem 3.1. Let $\Gamma \vdash \Delta$ be a LK^{++} -derivable sequent. By Proposition 3.6, its Skolemization $sk(\Gamma) \vdash sk(\Delta)$ is derivable. By Proposition 3.5, the standard translation of $sk(\Gamma) \vdash sk(\Delta)$ is LK^ε -derivable, whence, by Proposition 3.4, $sk(\Gamma) \vdash sk(\Delta)$ is already LK-derivable. Finally, it follows from [1, Theorem 2] that the unskolemized form $\Gamma \rightarrow \Delta$ is also LK-derivable. \square

4 Non-elementary speed-up

Our strategy for proving Theorem 2.6 is to show that LK^+ simulates a strong calculus that is already non-elementarily faster than LK. The main feature of this calculus is the addition of certain rules for quickly shifting quantifiers.

Definition 4.1. The calculus LK_{shift} is obtained by extending LK with the following rules:

$$\frac{\Gamma, \kappa[Qx A \triangleleft B] \vdash \Delta}{\Gamma, \kappa[Q'x (A \triangleleft B)] \vdash \Delta} \quad \frac{\Gamma, \kappa[A \triangleleft Qx B] \vdash \Delta}{\Gamma, \kappa[Q'x (A \triangleleft B)] \vdash \Delta}$$

$$\frac{\Gamma \vdash \Delta, \kappa[Qx A \triangleleft B]}{\Gamma \vdash \Delta, \kappa[Q'x (A \triangleleft B)]} \quad \frac{\Gamma \vdash \Delta, \kappa[A \triangleleft Qx B]}{\Gamma \vdash \Delta, \kappa[Q'x (A \triangleleft B)]}$$

where $\kappa[\cdot]$ is a context, $\triangleleft \in \{\wedge, \vee, \rightarrow\}$ and $Q' = Q$, except if \triangleleft is \rightarrow and Q is taken from the antecedent, in which case Q' is opposite. We refer to these rules as *quantifier shifts*.

Proposition 4.2. *LK^+ simulates LK_{shift} exponentially, i.e., every LK_{shift} -provable sequent is LK^+ -provable and there is an exponential function that bounds the length of the least cut-free LK^+ -proof of a LK^+ -provable sequent in terms of its least cut-free LK_{shift} -proof.*

Proof. Let π be a cut-free LK_{shift} -proof of size n . We transform it into a cut-free LK^+ -proof. Assume by induction that there is only one application of a quantifier shift and this is the last inference. We proceed by cases. For simplicity, we assume κ does not change the polarity of the quantifiers. We also assume \triangleleft is \rightarrow ; the other connectives are treated similarly.

CASE I. The last inference is:

$$\frac{\Gamma, \kappa[\forall x A(x) \rightarrow B] \vdash \Delta}{\Gamma, \kappa[\exists x (A(x) \rightarrow B)] \vdash \Delta}$$

so that the proof has the following structure:

$$\frac{\frac{\frac{\vdots \sigma_2}{\Gamma'' \vdash \Delta'', A(a)}}{\Gamma'' \vdash \Delta'', \forall x A(x)} (*)}{\frac{\frac{\vdots \sigma_1}{\Gamma' \vdash \Delta', \forall x A(x)} \quad \frac{\vdots \sigma_3}{\Gamma', B \vdash \Delta'}}{\Gamma', \forall x A(x) \rightarrow B \vdash \Delta'}} \frac{\vdots \sigma_0}{\Gamma, \kappa[\forall x A(x) \rightarrow B] \vdash \Delta} \frac{\vdots \sigma_0}{\Gamma, \kappa[\exists x (A(x) \rightarrow B)] \vdash \Delta}$$

where the σ_i denote subproofs. We modify the proof. Our strategy is as follows: we would like to merge the subproofs σ_1 and σ_2 by simply postponing the inference $(*)$ as follows:

$$\frac{\frac{\frac{\vdots \sigma_2 + \sigma_1}{\Gamma' \vdash \Delta', A(a)}}{\Gamma', A(a) \rightarrow B \vdash \Delta'} \quad \frac{\vdots \sigma_3}{\Gamma', B \vdash \Delta'}}{\Gamma', \exists x (A(x) \rightarrow B) \vdash \Delta'} (**)$$

$$\frac{\vdots \sigma_0}{\Gamma, \kappa[\exists x (A(x) \rightarrow B)] \vdash \Delta}$$

The problem that might arise is that some occurrence of $\forall x A(x)$ that would be contracted in σ_1 with the indicated occurrence is unable to be contracted. Assume for simplicity there is only one such occurrence. To solve this, notice that the problematic occurrence of $\forall x A(x)$ must originate from a strong-quantifier inference. Omit that inference and drag the unquantified formula $A(b)$ until after $(**)$, so that at that point we have a derivation of

$$\Gamma'_0, A(b), \Gamma'_1, \exists x (A(x) \rightarrow B), \vdash \Delta',$$

where $\Gamma' = \Gamma'_0, \Gamma'_1$. Add to that some exchanges to obtain:

$$\Gamma', \exists x (A(x) \rightarrow B), A(b) \vdash \Delta',$$

and a subproof

$$\frac{\vdots \sigma_3}{\Gamma', B \vdash \Delta'} \quad (9)$$

in order to infer

$$\Gamma', \exists x (A(x) \rightarrow B), A(b) \rightarrow B \vdash \Delta'.$$

with an application of $\exists L$ and a contraction, we are left again with

$$\Gamma', \exists x (A(x) \rightarrow B) \vdash \Delta',$$

to which we can apply σ_0 as desired. The proof grows by adding a copy of (9)—of size at most n —for each problematic occurrence of $\forall x A(x)$ (of which there are at most n). Hence, it grows quadratically.

CASE II. The last inference is:

$$\frac{\Gamma, \kappa[\exists x A(x) \rightarrow B] \vdash \Delta}{\Gamma, \kappa[\forall x (A(x) \rightarrow B)] \vdash \Delta}$$

The proof has the following structure:

$$\frac{\frac{\frac{\vdots \sigma_2}{\Gamma'' \vdash \Delta'', A(t)}}{\Gamma'' \vdash \Delta'', \exists x A(x)}}{\frac{\frac{\vdots \sigma_1}{\Gamma' \vdash \Delta', \exists x A(x)} \quad \frac{\vdots \sigma_3}{\Gamma', B \vdash \Delta'}}{\Gamma', \exists x A(x) \rightarrow B \vdash \Delta'}} \quad \frac{\vdots \sigma_0}{\frac{\Gamma, \kappa[\exists x A(x) \rightarrow B] \vdash \Delta}{\Gamma, \kappa[\forall x (A(x) \rightarrow B)] \vdash \Delta}}$$

We would like to merge the subproofs σ_1 and σ_2 :

$$\frac{\frac{\frac{\vdots \sigma_2 + \sigma_1}{\Gamma' \vdash \Delta', A(t)}}{\Gamma', A(t) \rightarrow B \vdash \Delta'}}{\frac{\Gamma', \forall x (A(x) \rightarrow B) \vdash \Delta'}{\frac{\vdots \sigma_0}{\Gamma, \kappa[\forall x (A(x) \rightarrow B)] \vdash \Delta}}}$$

As before, we face the problem of circumventing a contraction of $\exists x A(x)$ in σ_1 and solve it the same way. The proof grows quadratically again. We note the following: it might happen that $t = a$ for some free variable a . Hence, this proof does not go through for LK, as a could be the characteristic variable of a strong-quantifier inference in σ_1 .

CASE III. The last inference is:

$$\frac{\Gamma, \kappa[A \rightarrow \forall x B(x)] \vdash \Delta}{\Gamma, \kappa[\forall x (A \rightarrow B(x))] \vdash \Delta}$$

This is analogous to CASE II.

CASE IV. The last inference is:

$$\frac{\Gamma, \kappa[A \rightarrow \exists x B(x)] \vdash \Delta}{\Gamma, \kappa[\exists x (A \rightarrow B(x))] \vdash \Delta}$$

This is analogous to CASE I.

CASE V. The last inference is:

$$\frac{\Gamma \vdash \Delta, \kappa[\forall x A(x) \rightarrow B]}{\Gamma \vdash \Delta, \kappa[\exists x (A(x) \rightarrow B)]}$$

The proof has the following form:

$$\frac{\frac{\frac{\vdots \sigma_2}{\Gamma'', A(t) \vdash \Delta'', B}}{\Gamma'', \forall x A(x) \vdash \Delta'', B} (*)}{\frac{\frac{\vdots \sigma_1}{\Gamma', \forall x A(x) \vdash \Delta', B}}{\Gamma' \vdash \Delta', \forall x A(x) \rightarrow B}} \frac{\frac{\vdots \sigma_0}{\Gamma \vdash \Delta, \kappa[\forall x A(x) \rightarrow B]}}{\Gamma \vdash \Delta, \kappa[\exists x (A(x) \rightarrow B)]}$$

As before, we would like to postpone the inference $(*)$, thus merging σ_1 and σ_2 :

$$\frac{\frac{\frac{\vdots \sigma_2 + \sigma_1}{\Gamma', A(t) \vdash \Delta', B}}{\Gamma' \vdash \Delta', A(t) \rightarrow B} (**)}{\frac{\Gamma' \vdash \Delta', \exists x (A(x) \rightarrow B)}{\Gamma \vdash \Delta, \kappa[\exists x (A(x) \rightarrow B)]} \frac{\vdots \sigma_0}{\Gamma \vdash \Delta, \kappa[\exists x (A(x) \rightarrow B)]}}$$

However, we might have—similarly to the previous cases—an occurrence of $\forall x A(x)$ that would be contracted on the left-hand side as part of σ_1 . For simplicity, assume there is only one. This occurrence is inferred from a formula $A(s)$. Postpone this occurrence until after $(**)$ so that at that point we have a derivation of

$$\Gamma'_0, A(s), \Gamma'_1 \vdash \Delta', \exists x (A(x) \rightarrow B),$$

where $\Gamma' = \Gamma'_0, \Gamma'_1$. Add to that some exchanges to obtain:

$$\Gamma', A(s) \vdash \Delta', \exists x (A(x) \rightarrow B),$$

and a weakening:

$$\frac{\Gamma', A(s) \vdash \Delta', \exists x (A(x) \rightarrow B)}{\Gamma', A(s) \vdash \Delta', \exists x (A(x) \rightarrow B), B} WR \quad (10)$$

in order to infer

$$\Gamma' \vdash \Delta', \exists x (A(x) \rightarrow B), A(s) \rightarrow B.$$

with an application of $\exists R$ and a contraction, we are left again with

$$\Gamma' \vdash \Delta', \exists x (A(x) \rightarrow B),$$

to which we can apply the rest of σ_0 as desired. In this case the proof only grows linearly. As in CASE II, this proof would not go through for LK.

CASE VI. The last inference is:

$$\frac{\Gamma \vdash \Delta, \kappa[\exists x A(x) \rightarrow B]}{\Gamma \vdash \Delta, \kappa[\forall x (A(x) \rightarrow B)]}$$

The proof has the following form:

$$\frac{\frac{\frac{\frac{\vdots \sigma_2}{\Gamma'', A(a) \vdash \Delta'', B}}{\Gamma'', \exists x A(x) \vdash \Delta'', B} (*)}{\Gamma', \exists x A(x) \vdash \Delta', B} \vdots \sigma_1}{\Gamma' \vdash \Delta', \exists x A(x) \rightarrow B} \vdots \sigma_0$$

$$\frac{\Gamma \vdash \Delta, \kappa[\exists x A(x) \rightarrow B]}{\Gamma \vdash \Delta, \kappa[\forall x (A(x) \rightarrow B)]} \vdots \sigma_0$$

As before, we would like to postpone the inference $(*)$, thus merging σ_1 and σ_2 :

$$\begin{array}{c}
\frac{\dot{\vdash} \sigma_2 + \sigma_1}{\Gamma', A(a) \vdash \Delta', B} \\
\hline
\frac{\Gamma' \vdash \Delta', A(a) \rightarrow B}{\Gamma' \vdash \Delta', \forall x (A(x) \rightarrow B)} \\
\hline
\frac{\dot{\vdash} \sigma_0}{\Gamma \vdash \Delta, \kappa[\forall x (A(x) \rightarrow B)]}
\end{array}$$

We deal with contractions of $\exists x A(x)$ on the left-hand side as in the previous case. The proof grows linearly.

CASE VII. The last inference is:

$$\frac{\Gamma \vdash \Delta, \kappa[A \rightarrow \forall x B(x)]}{\Gamma \vdash \Delta, \kappa[\forall x (A \rightarrow B(x))]}$$

This is analogous to CASE VI.

CASE VIII. The last inference is:

$$\frac{\Gamma \vdash \Delta, \kappa[A \rightarrow \exists x B(x)]}{\Gamma \vdash \Delta, \kappa[\exists x (A \rightarrow B(x))]}$$

This is analogous to CASE V.

Hence, we have dealt with all the cases. Note that for each quantifier shifted, the proof grows at most quadratically. Since there are at most n quantifier shifts, the size of the resulting proof is bounded by

$$n^{2^n} = 2^{2^n \cdot \log(n)} \approx 2^{2^n}.$$

This finishes the proof. \square

Theorem 2.6 is a consequence of Proposition 4.2 and the following result:

Theorem 4.3. *There is no elementary function bounding the length of the shortest cut-free LK-proof of a formula in terms of its shortest cut-free LK_{shift} -proof.*

Proof. We will make use of a very specific family of sequents $\{S_i\}_{i < \omega}$ described in [2] and due to Statman [7], and specific LK-proofs thereof. The sequents and the proofs themselves are not important for our proof. What is relevant is that they have the following properties:

1. the size of S_i is polynomial in i ;
2. there is no bound on the size of their smallest cut-free LK-proofs that is elementary on i ;

3. the size of these proofs (with cuts), however, is polynomially bounded on i ;
4. all cut formulae are closed; we can also assume they are prenex by, e.g., [3, Theorem 3.3].

Let $\Gamma_i \vdash \Delta_i$ be one of the sequents. We modify the proof as follows: first, replace each cut

$$\frac{\Gamma \vdash \Delta, A \quad A, \Gamma \vdash \Delta}{\Gamma \vdash \Delta} \text{Cut}$$

with an application of $\rightarrow L$:

$$\frac{\Gamma \vdash \Delta, A \quad \Gamma, A \vdash \Delta}{\Gamma, A \rightarrow A \vdash \Delta} \rightarrow L$$

We are left with a cut free proof π_0 whose end sequent is of the form:

$$A_0 \rightarrow A_0, \dots, A_m \rightarrow A_m, \Gamma_i \vdash \Delta_i. \quad (11)$$

Choose any occurrence of a quantifier in A_0 that is not in the scope of another quantifier. Since it appears once in the antecedent of $A_0 \rightarrow A_0$ and once in the consequent, it appears once with each polarity. Because A_0 is assumed to be prenex, we can apply two quantifier shifts, so that we are left with a proof π'_0 of the sequent:

$$\forall x_0^0 \exists x_1^0 (A'_0 \rightarrow A'_0), \dots, A_m \rightarrow A_m, \Gamma_i \vdash \Delta_i, \quad (12)$$

Continue choosing any occurrence of a quantifier in A'_0 that is not in the scope of another quantifier (within A'_0) and applying quantifier shifts in pairs until we obtain a proof π_1 of the sequent:

$$\forall x_0^0 \exists x_1^0 \dots (\hat{A}_0 \rightarrow \hat{A}_0), \dots, A_m \rightarrow A_m, \Gamma_i \vdash \Delta_i,$$

where \hat{A}_0 is quantifier-free and the prefix of $\hat{A}_0 \rightarrow \hat{A}_0$ consists of alternating quantifiers. Repeat this procedure for each of the A_i to obtain a proof π_m where each formula $A_i \rightarrow A_i$ is replaced by an expression of the form:

$$\forall x_0^i \exists x_1^i \dots (\hat{A}_i \rightarrow \hat{A}_i),$$

where \hat{A}_i is quantifier free and the quantifier prefix is alternating. Let $\hat{\Gamma}^i \vdash \hat{\Delta}^i$ be the sequent:

$$\forall x_0^0 \exists x_1^0 \dots (\hat{A}_0 \rightarrow \hat{A}_0), \dots, \forall x_0^m \exists x_1^m \dots (\hat{A}_m \rightarrow \hat{A}_m), \Gamma_i \vdash \Delta_i. \quad (13)$$

Note that, since the size of the initial proof was polynomial in i , there were polynomially many quantifiers in the proof; hence, we added only polynomially many quantifier shifts, and so the size of the resulting proof π_m of (13) is bounded polynomially in i . Moreover, it is cut-free. Consequently, it suffices to show:

Claim 4.4. *There is no elementary function bounding the size of the smallest cut-free LK-proofs of (13).*

Proof. Let $\{\sigma_i\}_{i < \omega}$ be a sequence of such proofs. First, we transform it into a sequence of proofs of the Skolemizations of the sequents (13), so that for each sequent, each of the implications $\hat{A}_i \rightarrow \hat{A}_i$ remains only universally quantified. Then, by Herbrand's theorem, there are propositional proofs $\{\theta_i\}_{i < \omega}$ of the Herbrand sequents of (13), each of which is of the form:

$$\bigwedge_k B_k^0, \dots, \bigwedge_k B_k^m, \Gamma' \vdash \Delta'. \quad (14)$$

Moreover, the lengths of the Herbrand sequents are bounded exponentially by the lengths of $\{\sigma_i\}_{i < \omega}$ (see [2, Theorem 4.3]). Each conjunct B_k^p is a quantifier-free implication of the form:

$$A(t_1, \dots, t_l) \rightarrow A(s_1, \dots, s_l), \quad (15)$$

for some terms $t_1, \dots, t_l, s_1, \dots, s_l$. The key point is that, in the process of Skolemizing the sequents (13), exactly one term in each pair (t_j, s_j) was weakly quantified and the other (which was under the scope of the strong quantifier) was replaced by a Skolem function. Hence, either t_j is of the form $f(s_j, \vec{r})$ or s_j is of the form $f(t_j, \vec{r})$.

Transform the sequent (14) as follows: pick a pair (t_j, s_j) of some disjunct B_k^p such that the element thereof that was replaced by a Skolem term—say, $f(s_j, \vec{r}) = t_j$ —does not have any other of $t_1, \dots, t_l, s_1, \dots, s_l$ as an argument. Such a term of course exists—it is the first term whose quantifier was shifted in (12). Substitute s_j for t_j throughout the sequent.

By repeating this process sufficiently-many times, each formula (15) is transformed into a propositional tautology of the form $A(\vec{t}) \rightarrow A(\vec{t})$. This means that the sequent

$$\vdash \bigwedge_k B_k^0, \dots, \bigwedge_k B_k^m.$$

is also a propositional tautology, whereby so too is

$$\Gamma' \vdash \Delta',$$

from which follows that it is LK-provable. By [1, Theorem 2], there is an LK-proof of the unskolemized sequent of length exponential in that of $\Gamma' \vdash \Delta'$. But this is impossible, since—by assumption—there exist no short proofs of the unskolemized sequent. \square

This establishes Theorem 4.3 and, hence, Theorem 2.6. \square

5 Consequences of the Speed-up Theorem

We describe some consequences of Theorem 2.6:

Corollary 5.1.

1. *There is no elementary bound on the size of the smallest cut-free LK-proof of the Skolemization of a sequent in terms of its smallest cut-free LK⁺-proof (resp. LK⁺⁺-proof).*
2. *There is no elementary bound on the length of the Herbrand sequent of a sequent in terms of its smallest cut-free LK⁺-proof (resp. LK⁺⁺-proof).*

Proof. The first item implies the second, as a cut-free LK⁺⁺-proof of the Skolemization of a sequent contains no strong quantifier and is hence already an LK-proof, whence we can apply Herbrand's theorem.

To see that the first item holds, notice that if we had such a bound, we could argue as for Claim 4.4 to obtain a contradiction. \square

Definition 5.2. We say that a cut-eliminating procedure is *Gentzen-style* if it is a transformation of proofs consisting of permutation of rules, substitution of free-variables, and absorption of axioms, i.e., elimination of cuts

$$\frac{\Gamma \vdash \Delta, A \quad A \vdash A}{\Gamma \vdash \Delta, A} \text{ Cut}$$

by deleting the indicated occurrence of $A \vdash A$.

An important consequence of cut elimination is the *subformula property*—every derivable sequent S has a cut-free derivation wherein all formulae are subformulae of S . Consider LJ⁺, the analog of the extended calculus LK⁺ for intuitionistic logic, namely, the calculus obtained by restricting ourselves to sequents with exactly one formula on the right-hand side. This calculus is not sound for intuitionistic logic, as shown by any of the examples in the introduction. A consequence of this is the following:

Proposition 5.3. LK⁺ and LK⁺⁺ admit no Gentzen-style cut elimination.

Proof. Consider example (2). Since only one formula appears on the right-hand side of each sequent, this is an LJ⁺-proof. If there were a Gentzen-style cut elimination for LK⁺, we could apply it to the example (2) to obtain a cut-free proof of

$$\vdash \exists x (A(x) \rightarrow A(f(x))). \quad (16)$$

However, by the subformula property, this derivation would consist entirely of subformulae of (16). In particular, it would contain no strong quantifier inferences. Therefore, it would already be an LJ-proof. But this is impossible, as (16) is not intuitionistically valid. \square

6 Concluding Remarks

Many related questions remain open. The most important question consists of fully investigating the role that inferences play in proofs, both from a philosophical and from a technical standpoint. We believe that it is not straightforward, as the present work provides evidence for challenging the traditional view of inferences simply as one-step subproofs of proofs, or—conversely—of proofs simply as arbitrary concatenations of inferences.

Another problem is to explore the addition of different unsound rules to proof systems. Here, we gave two examples, but many more are possible. It is also not clear exactly how faster the calculus LK^+ really is—the weakening of the usual characteristic-variable condition into substitutability, weak regularity, and the side-variable condition can perhaps be exploited further.

It is not clear whether LK^+ is any (or significantly) slower than LK^{++} . It is also not clear in what precise relation they both stand to LK_{shift} and LK^ε . All these questions seem to promise fruitful lines of future research.

References

- [1] M. Baaz, S. Hetzl, and D. Weller. On the complexity of proof deskolemization. *The Journal of Symbolic Logic*, 77:669–686, 2012.
- [2] M. Baaz and A. Leitsch. On skolemization and proof complexity. *Fundamenta Informaticae*, 20:353–379, 1994.
- [3] M. Baaz and A. Leitsch. Cut normal forms and proof complexity. *Annals of Pure and Applied Logic*, 97:127–177, 1999.
- [4] D. Hilbert and P. Bernays. *Grundlagen der Mathematik II*. Springer-Verlag, 1970.
- [5] G. Moser and R. Zach. The Epsilon Calculus and Herbrand Complexity. *Studia Logica*, 82(1):133–155, 2006.
- [6] V. P. Orevkov. Lower Bounds for Increasing Complexity of Derivations after Cut Elimination (in Russian). *Journal of Soviet Mathematics*, pages 2337–2350, 1982.
- [7] R. Statman. Lower Bounds on Herbrand’s Theorem. *Proceedings of the American Mathematical Society*, 75:104–107, 1979.